ESTIMATION OF OBESITY LEVELS BASED ON EATING HABITS AND PHYSICAL CONDITIONS

Abstract

This study aims to estimate obesity levels by examining the relationships between eating habits, physical activity, and the likelihood of obesity. A regression model is developed using a comprehensive dataset to predict obesity based on key variables.

Dataset

This dataset includes information for predicting obesity levels in individuals from Mexico, Peru, and Colombia, considering their eating habits and physical condition. It consists of 17 attributes and encompasses a total of 2111 records.

The dataset was taken from UCI Machine Learning Repository;

https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition
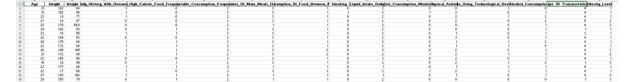
Methodology

This study utilizes a dataset for estimating obesity levels in individuals from Mexico, Peru, and Colombia based on their eating habits and physical condition. The initial step involves cleaning and preprocessing the data, addressing aspects such as column names, data types, and converting categorical values to numerical ones. Subsequently, data visualization techniques are applied to uncover patterns and relationships within the dataset. Statistical methods, including the use of correlation matrices, are then employed to identify relevant features. To construct a regression model, the dataset is divided into a training set and a test set. Standard scaling is

implemented to normalize the data. Following model construction, an evaluation of its performance is conducted to assess its effectiveness in estimating obesity levels.

Before Preprocessing

After Preprocessing

Model Development

Building a regression model;

Building And Fitting Regression model

```
In [13]:  # Building Multiple Linear Regression model
          from sklearn.linear_model import LinearRegression
          model = LinearRegression()
```

```
In [14]:  # Fitting Multiple Linear Regression to the Training set
          model.fit (X_train_scaled, y_train)
```

Out[14]: LinearRegression()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

Model Evaluation

```
In [19]:  # Calculating the R squared value
          from sklearn.metrics import r2_score
          r2_score(y_test, y_pred)
```

```
Out[19]:  0.8701434870082194
```

```
In [20]:  import seaborn as sns
          sns.displot(y_pred-y_test,kind='kde')
```

```
Out[20]:  <seaborn.axisgrid.FacetGrid at 0x152953d30d0>
```